

How Should Financial Intermediation Services be Taxed?*

Ben Lockwood[†]

First version: May 31 2010

This version: April 17 2012

Abstract

This paper considers the optimal taxation of savings intermediation and payment services in a dynamic general equilibrium setting, when the government can also use consumption and income taxes. When payment services are used in strict proportion to final consumption, and the cost of intermediation services is fixed and the same across firms, the optimal taxes are generally indeterminate. But, when firms differ exogenously in the cost of intermediation services, the tax on savings intermediation should be zero. Also, when household time and payment services are substitutes in transactions, the optimal tax rate on payment services is determined by the returns to scale in the conditional demand for payment services, and is generally different to the optimal rate on consumption goods. The extension to the case of n commodities is studied, and conditions sufficient for uniform taxation of goods and payment services within a period are obtained, generalizing the analysis of Auerbach and Gordon(2002).

JEL Classification: G21, H21, H25

Keywords: financial intermediation services, tax design, banks, monitoring, payment services

*I would like to thank Steve Bond, Clemens Fuest, Michael Devereux, Michael McMahon, Miltos Makris, Ruud de Mooij, and seminar participants at the University of Southampton, GREQAM, the 2010 CBT Summer Symposium, and the 2011 IIPF Conference for helpful comments on an earlier draft. I also gratefully acknowledge support from the ESRC grant RES-060-25-0033, "Business, Tax and Welfare".

[†]CBT, CEPR and Department of Economics, University of Warwick, Coventry CV4 7AL, England; Email: B.Lockwood@warwick.ac.uk

1. Introduction

Financial intermediation services include such important services as intermediation between borrowers and lenders, insurance, and payment services (e.g. credit and debit card services). These services comprise a significant and growing part of the national economy; for example, financial intermediation services, measured using the OECD methodology¹, were 3.9% of GDP in the UK in 1970, and increased to 7.9% by 2005. The figures for the Eurozone countries as a whole are 2.7% to 5.5%. In the US, the finance and insurance sector, excluding real estate, which includes financial intermediation, accounted for 7.3% of US value-added in 1999, rising to 8.4% in 2009².

The question of whether, and how, financial intermediation services should be taxed is a contentious one. In the tax policy literature, it is largely assumed that within a consumption tax system, such as a VAT, it is desirable to tax financial services. For example, the European Commission has recently proposed changes to the VAT treatment of financial services within the European Union, so as bring these more within the scope of VAT (de la Feria and Lockwood (2010)). Also, the recent IMF proposals for a "bank tax" to cover the cost of government interventions in the banking system include a Financial Activities Tax levied on bank profits and remuneration, which would work very much like a VAT, levied using the addition method (IMF(2010)).

But, it is also recognized that there are technical difficulties in taxing financial intermediation when those services are not explicitly priced (so-called margin-based services), such as the intermediation between borrowers and lenders. This raises a problem for the use of a VAT via the usual invoice-credit method, for example (Ebril, Keen, Bodin and Summers(2001)). As a result of this, the status quo in most countries is that a wide range of financial intermediation services are not taxed³. However, conceptually, the problems can be solved, for example, by use of a cash-flow VAT (Hoffman et. al.(1987), Poddar and English(1997), Huizinga(2002), Zee(2005)), and the increasing sophistication of banks' IT systems means that these solutions are also becoming practical.

So, it is increasingly relevant to ask, setting aside the technical problems, should financial intermediation services supplied to households be taxed at all? And if so, at

¹See <http://www.euklems.net>.

²See http://www.bea.gov/industry/gdpbyind_data.htm

³For example, in the EU, the Sixth VAT Directive and subsequent legislation exempts a wide range of financial services from VAT, including insurance and reinsurance transactions, the granting and the negotiation of credit, transactions concerning deposit and current accounts, payments, transfers, debts, cheques, currency, bank notes and coins used as legal tender etc. (Council Directive 2006/112/EC of 28 November 2006, Article 135).

what rates? Given the overall importance of financial services to modern economies, there is surprisingly little written on this more fundamental question (Section 2 has a discussion of the literature). Moreover, we would argue that the existing literature does not really clarify which of the fundamental principles of tax design apply. For example, is it the case that financial intermediation services are intermediate goods in the production of final consumption for households, and thus should not be taxed? Or, should they be taxed at the same rate as other goods purchased on the market, at least under conditions when a uniform consumption tax is optimal?

The objective of this paper is to address these fundamental questions⁴. We set up and solve the tax design problem in a dynamic general equilibrium model of the Chamley(1986) type, where the government chooses taxes on payment services and savings intermediation, as well as the usual taxes on consumption (or equivalently, wage income) and income from capital, and where financial intermediaries, in the form of banks, are explicitly modelled. On the payment services side, we assume, following the literature on the transactions cost approach to the demand for money, that payment services are not necessarily proportional to consumption, but can be used to economize on the household time input to trading. This is realistic: for example, making use of a basic bank account requires a time input, e.g. trips to the bank, but use of an additional payment service e.g. a credit card, substitutes for trips to the bank.

We assume that the cost of savings intermediation per unit of capital is fixed, but can vary across borrowers (firms). Again, this is realistic; savings intermediation is a complex process involving initial assessment of the borrower via e.g. credit scoring, structuring and pricing the loan, and monitoring compliance with loan covenants (Gup and Kolari(2005, chapter 9). The extension to variable costs of savings intermediation is addressed in Lockwood(2010).

We then solve the tax design problem, where the government has access to a full set of taxes, i.e. the usual wage and capital income taxes, plus a tax on the consumption good and on payment services, and a tax paid by the bank on the spread between borrowing and lending rates. We set the wage tax equal to zero to eliminate the usual tax indeterminacy via the household budget constraint, and focus on the four remaining taxes.

The tax on savings intermediation is determined as follows. In the tax design problem, the tax on capital income is used as the instrument to pin down the rate of substitution

⁴It should be noted that this paper does *not* deal with corrective taxes on bank lending designed to internalize the social costs of bank failure or the costs of bailout; on this, see e.g. Hellmann, Murdock and Stiglitz(2000), Keen(2010) or Bianchi and Mendoza(2010).

between present and future consumption for the household. So, this means that the tax on savings intermediation is a "free instrument" that can be used to ensure that capital is allocated efficiently across firms. In turn, the cost of capital to a particular firm will be the cost of capital to the bank i.e. the return paid to depositors, plus the cost of intermediation, where the latter includes any tax. A non-zero tax on savings intermediation will distort the relative cost of capital across firms, and so this tax is optimally set to zero. This is a version of the Diamond-Mirrlees production efficiency result.

Turning to the tax on the payment service, our first result is that the total tax "wedge" between consumption and leisure is a weighted average of the tax on consumption and on the payment service, and is determined by a standard optimal tax formula, involving the general equilibrium expenditure elasticity of consumption (Atkeson, Chari, and Kehoe(1999)). However, the sign of the tax on the payment service itself is determined⁵ *not* by the structure of preferences, but by the properties of the conditional demand for payment services as a function of the household consumption level and time input to transactions, or "shopping time". In particular, when this conditional demand has constant returns with respect to these variables, payment services should be untaxed; this can be understood as an instance of the Diamond-Mirrlees production efficiency result. The general conclusion is that the tax on payment services is determined in a completely different way to the tax on consumption, and thus will in general be at a different rate.

We then extend the analysis to many consumption goods in each period. We make the plausible assumption that there is only one kind of payment service that must be taxed at the same rate, whichever good it is used to purchase. We show that the presence of payment services generally makes it less likely that commodity taxation will be uniform within the period. Moreover, *even if* conditions hold for commodity taxation to be uniform, this does not imply that payment services need to be taxed at the same uniform rate as commodities. Generally, what is required for the same uniform tax on *both* goods and payment services are: (i) standard conditions for uniformity i.e. separability in goods and leisure, plus a homothetic goods sub-utility function, *plus* either (ii) no time input to transactions, or (iii) that time and payment services are not substitutable in the production of transactions services i.e. Leontief technology, *and* the ratio of payment services to time required is uniform across goods. These results generalize, and demonstrate the limitations of, the Auerbach-Gordon(2002) result that if goods are subject to a uniform

⁵Strictly speaking, this requires the conditional demand for payment services to be Cobb-Douglas, but it is also likely to hold for a variety of other cases, see Section 4.

tax, payment services should be subject to the same tax.

The remainder of the paper is organized as follows. Section 2 discusses related literature. Section 3 outlines the model, and explains how existing contributions can be viewed as special cases. Section 4 presents the main results. Section 5 studies the case of n commodities, Section 6 considers other extensions, and Section 7 concludes.

2. Related Literature

There is a small literature directly addressing the optimal taxation of borrower-lender intermediation and payment services, Grubert and Mackie(1999), Jack(1999), Auerbach and Gordon (2002), and Boadway and Keen(2003). Using for the most part a simple two-period consumption-savings model, these papers broadly agree on a policy prescription⁶. Given a consumption tax that is uniform across goods (at a point in time), payment services should be taxed at this uniform rate, but savings intermediation should be left untaxed. The argument used to establish this is simple; in a two-period consumption-savings model with the same, exogenously fixed, tax on consumption in both periods, this arrangement leaves the marginal rate of substitution between current and future consumption undistorted i.e. equal to the marginal rate of transformation⁷.

However, one can make three criticisms of the current literature. First, even taking their set-up as given, their optimal taxes are indeterminate. Purely mathematically, two taxes cannot be uniquely determined from a single efficiency condition. Second, in their analysis, consumption (wage) and capital income taxes are taken as given, and not optimized by the government. Third, relative to the model of this paper, the models analyzed in the current literature are very special in a number of respects. For example, implicitly, these papers are assuming⁸ a fixed labour supply, so that a uniform tax on consumption over the life-cycle is first-best efficient, as it does not distort the inter-temporal allocation of consumption, and thus financial intermediation should not do so

⁶Chia and Whalley(1999), using a computational approach, reach the rather different conclusion that no intermediation services should be taxed, but but their model is not directly comparable to these others, as the intermediation costs are assumed to be proportional to the *price* of the goods being transacted.

⁷Auerbach and Gordon(2002) have a model that is in some respects more general, and they also take a different analytical approach. Specifically, their model allows for T periods, multiple consumption goods, and variable labour supply. In this setting, they show that a uniform tax on all commodities and payment services is equivalent to a wage tax. Thus, they show that *if* a uniform commodity tax is optimal, payment services should be taxed at the same rate, consistently with the other literature cited.

⁸The exception here is Auerbach and Gordon(2002), where labour supply is variable. However, in their model, the consumption tax is just assumed to be uniform, not optimised.

either. Again, special assumptions are made about the demand for payment services, and intermediation activities of banks. Specifically, they assume (i) that payment services are consumed in proportion to consumption; (ii) that the costs of savings intermediation are in proportion to capital invested. We are able to show that the basic result of this literature - i.e. that intermediation taxes are indeterminate, but that *an* optimal tax structure is to tax payment services at the same rate as consumption, but exempt savings intermediation - also emerges in our model when all of these special assumptions are made (Proposition 1 below).

A less closely related literature is that on the optimal inflation tax which takes a transaction costs approach to the demand for money (Kimbrough(1986), Guidotti and Vegh(1993), Correia and Teles(1996, 1999)). In this literature, money formally plays a role similar to payment services in our model; the main differences are (i) that it is assumed a free good i.e. it has a zero production cost, and (ii) it is subject to an inflation tax, rather than a fiscal tax. While (ii) makes no difference from an analytical point of view, (i) does; it turns out that when money is free, the optimal inflation tax is zero, as long as the transactions demand for household time is a homogenous function of money and consumption. A much more closely related finding is in Correia and Teles (1996), where, in Section 3 of their paper, money is allowed to have a positive production cost. Proposition 3 below can be regarded as an extension of Proposition 2 in their paper.

Finally, one can interpret the use of payment services and time as inputs to household production of final consumption goods, so our analysis is linked to the small literature on optimal taxation with household production, particularly Kleven(2004). These links are explored further in Section 6.1 below.

3. The Model

3.1. Households

The model is a version of Atkeson, Chari and Kehoe(1999) with payment services and savings intermediation. There is a single infinitely lived household with preferences over levels of a single consumption good, leisure, and a public good in each period $t = 0, ..\infty$ of the form

$$\sum_{t=0}^{\infty} \beta^t (u(c_t, l_t) + v(g_t)) \quad (3.1)$$

where c_t is the level of final consumption in period t , l_t is the consumption of leisure, and g_t is public good provision. Utilities $u(c, l), v(g)$ are strictly increasing and strictly

concave in their arguments.

We take a transactions cost approach to the demand for payment services⁹, and suppose that consumption c_t incurs a transaction cost in terms of household time, and this cost is reduced by the use of payment services x_t . For example, making use of a bank account requires a time input, e.g. trips to the bank, but use of an additional payment service e.g. a credit card substitutes for trips to the bank. Then we have $h_t = h(c_t, x_t)$, where h is increasing in c_t , and decreasing in x_t .

To help intuition, consider a special case where consumption level c_t requires c_t separate transactions, and where a fraction f of these transactions are undertaken from a transactional account with a credit card, and the remainder, $1 - f$, from a transactional account without a credit card, where visits to the bank to withdraw cash are required. Also assume for simplicity that there is no interest penalty with a transactional account i.e. it pays the same as a savings account¹⁰. Then, the amount of payment services required is $x_t = X + afc_t$, where X is the fixed cost of maintaining a credit card account, and a is the amount of payment services needed per transaction. Also, the household time needed is $h_t = \frac{(1-f)c_t}{b}$, where b is the number of transactions financed by a single trip to the bank, assumed fixed. Eliminating f from between these two equations gives a linear time demand function

$$h(c_t, x_t) = \frac{1}{b} \left(\frac{X - x_t}{a} + c_t \right) \quad (3.2)$$

It turns out that for our purposes, it is convenient to describe the implicit relationship $h_t = h(c_t, x_t)$ between c_t , h_t , and x_t in terms of the *conditional demand for payment services*

$$x_t = \phi(c_t, h_t), \quad \phi_c > 0, \quad \phi_h \leq 0 \quad (3.3)$$

In the case of the linear time demand function, (3.2), ϕ takes the form

$$\phi(c_t, h_t) = X + ac_t - abh_t \quad (3.4)$$

Specification (3.4) is interesting not only because it has plausible microfoundations, but also because it nests the existing literature as a special case: this literature effectively assumes ϕ independent of h_t and linear homogenous in c_t , i.e. $x_t = Ac_t$. This is of course

⁹This is of course, analogous to the transactions cost theory of the demand for money.

¹⁰This is, to a first approximation, a reasonable assumption; transactions accounts do pay similar rates of interest to instant access savings accounts. It is also analytically very convenient: without it, a version of the Baumol-Tobin inventory demand theory would come into play, implying that h_t , conditional on c_t , would also depend on w_t, r_t , which are determined in general equilibrium, and this would greatly complicate the analysis.

a special case of (3.4) where there are no fixed costs ($X = 0$), $A = a$, and no account without a credit card. In the general case, following Guidotti and Vegh(1993), we will assume that ϕ is convex in its arguments; this ensures that the household problem is concave.

The household thus supplies labour to the market of amount

$$m_t = 1 - l_t - h_t \quad (3.5)$$

where the total endowment of time per period is set at unity. In each period t , the household also saves k_{t+1} in units of the consumption good, and deposits it with a bank, who can then lend it on to firms who can use it as an input to production in the next period, after which they must repay the loan to the bank, who then in turn repays the household. So, in this model, capital only lasts one period. Finally, the household has no profit income in any period: firms may generate pure profits (see Section 3.2 below), but these are taxed at 100%.

In any period t , household is assumed to pay ad valorem taxes τ_t^c, τ_t^x on c_t, x_t , and also pays proportional taxes on labour and capital income. Using the well-known fact that an uniform consumption tax (i.e. $\tau_t^c = \tau_t^x$) is equivalent to a wage tax, we assume w.l.o.g. that the wage tax is zero. We also assume for convenience that one unit of the consumption good can be transformed into one unit of payment services or one unit of the public good. Moreover, in equilibrium, payment services are priced at marginal cost (see Section 3.3 below). This fixes the relative pre-tax price of c_t, x_t and g_t at unity.

So, the present value budget constraint of the household is

$$\sum_{t=0}^{\infty} p_t(c_t(1 + \tau_t^c) + k_{t+1} + x_t(1 + \tau_t^x)) = \sum_{t=0}^{\infty} p_t(w_t(1 - \tau_t)m_t + (1 + \rho_t)k_t) \quad (3.6)$$

where p_t is the price of output in period t , ρ_t is the after-tax return on capital to the household, and w_t is the wage. We normalize by setting $p_0 = 1$ and assume for convenience that $k_0 = 0$ i.e. initial capital is zero¹¹. Finally, $\rho_t = (1 - \tau_t^r)r_t$, where r_t is the pre-tax return on capital, determined below, and τ_t^r is the capital income tax.

Substituting (3.3),(3.5) in (3.6) gives:

$$\sum_{t=0}^{\infty} p_t(c_t(1 + \tau_t^c) + k_{t+1} + \phi(c_t, h_t)(1 + \tau_t^x)) = \sum_{t=0}^{\infty} p_t(w_t(1 - l_t - h_t) + (1 + \rho_t)k_t) \quad (3.7)$$

¹¹This simplifies the implementability constraint, and does not change anything of substance (see Atkeson, Chari and Kehoe(1999)).

The first-order conditions for a maximum of (3.1) subject to (3.7) with respect to c_t , l_t , h_t , k_{t+1} respectively are:

$$\beta^t u_{ct} = \lambda p_t (1 + \tau_t^c + \phi_{ct} (1 + \tau_t^x)) \quad (3.8)$$

$$\beta^t u_{lt} = \lambda p_t w_t \quad (3.9)$$

$$-\phi_{ht} (1 + \tau_t^x) = w_t \quad (3.10)$$

$$p_t = (1 + \rho_{t+1}) p_{t+1} \quad (3.11)$$

where λ is the multiplier on (3.7), and we use (here and below) the notation that for any any function f and variables x_t , y_t , the partial derivative of f with respect to x_t is f_{x_t} , the cross-derivative is $f_{x_t y_t}$ etc. Note that using this notation, the consumer price of final consumption is $p_t (1 + \tau_t^c + \phi_{ct} (1 + \tau_t^x))$, a weighted sum of the prices facing the household of c_t , and x_t .

3.2. Firms

There are firms, $i = 1, \dots, n$ with produce the homogenous good in each period. Firm i produces output from labour and capital via the strictly concave, decreasing returns, production function $F^i(k_t^i, m_t^i)$, where k_t^i, m_t^i are capital and labour inputs. We assume that firms face decreasing returns, because for fixed policy, they will generally face different costs of capital, but the same wage, and so in this environment, with constant returns, only the one firm with the lowest unit cost would operate. This case is of limited interest because then, a spread tax cannot affect the relative cost of capital of different firms, which means in turn that there is no efficiency argument for setting the spread tax to zero.

These firms are assumed to be perfectly competitive. But, they cannot purchase capital directly from households, but must borrow from banks. Moreover, we suppose that firms may differ in intermediation costs, as described in more detail in Section 3.3 below. So, firms face differences in the cost of capital i.e. firm i must repay $1 + r_t^i$ per unit of capital borrowed from the bank. Thus, profit-maximization implies:

$$F_m^i(k_t^i, m_t^i) = w_t, \quad F_k^i(k_t^i, m_t^i) = 1 + r_t^i \quad (3.12)$$

And, in addition, the capital and labour market clearing conditions are:

$$\sum_{i=1}^n k_t^i = k_t, \quad \sum_{i=1}^n m_t^i = 1 - l_t - h_t \quad (3.13)$$

These conditions (3.12),(3.13) jointly determine w_t and r_t , given household savings and labour supply decisions.

3.3. Banks

Banks in this economy provide two possible services. First, they can provide payment services to the households i.e. supply x_t . Second, they can provide intermediation between households and firms. Banks can compete on price for both these activities (i.e. households see the banks as perfect substitutes, both with respect to payment and intermediation services). We also assume no economies of scope, and constant returns in the provision of both services, so that banks must break even on both services. Assuming w.l.o.g. that the marginal and average cost of payment services is 1 in units of the consumption good, the price of payment services will also be 1 in equilibrium.

The cost of intermediating one unit of savings between the household and firm i is s^i . Note that we take s^i as fixed, but possibly varying between firms, for reasons discussed in the introduction. We also suppose that "spread" i.e. the value of intermediation services provided by the bank, can be taxed at some rate $\hat{\tau}_t^s$. In turn, the value of intermediation services is measured by $r_t^i - r_t$, where r_t^i is the lending rate to firm i , and r_t is the rate paid to depositors. So, $\hat{\tau}_t^s$ is a tax on both intermediation services provided to households, and to firms¹². Then, as banks make zero profit on this activity, we must have

$$(1 - \hat{\tau}_t^s)(r_t^i - r_t) - s^i = 0, \quad i = 1, \dots, n \quad (3.14)$$

Then, from (3.14):

$$r_t^i = r_t + (1 + \tau_t^s)s^i, \quad \tau_t^s = \frac{\hat{\tau}_t^s}{1 - \hat{\tau}_t^s} \quad (3.15)$$

We refer to τ_t^s as the *spread tax* from now on.

3.4. Discussion

The above model provides a general framework which encompasses the specific models of taxation of financial services (Auerbach and Gordon(2002), Boadway and Keen(2003)), Jack(1999), Grubert and Mackie(1999)) that have been developed so far. For example, Boadway and Keen(2003), Jack(1999), Grubert and Mackie(1999) are two-period versions of the above model¹³, with (implicitly) fixed labour supply. Auerbach and Gordon(2002)

¹²In principle, one could allow for the intermediation services received by these two parties to be taxed at different rates, but in practice, this is very difficult to implement (Poddar and English(1997)). Moreover, in our framework, the addition of this feature in our model would lead to tax indeterminacy.

¹³A minor qualification here is that Boadway and Keen allow for a fixed cost of savings intermediation e.g. fixed costs of opening a savings account. These introduce a non-convexity into household decision-making, which greatly complicates the optimal tax problem, and so we abstract from these in this paper.

is a finite-horizon version of the model, with the additional feature¹⁴ that there are n consumption goods in each period. This raises some new issues, and the case of n goods is covered in Section 5 below.

As already noted in Section 2, the feature of all these contributions, however, is the special assumptions they implicitly make about demand for payment services and bank intermediation. On the household side, they all assume, first, that payment services are needed *in fixed proportion to consumption* and that (implicitly) that a time input h_t is not required from the household. In our model, this amounts to the assumptions that $\phi(c_t, h_t) = Ac_t$ in (3.3), in which case, choosing the constant to be unity, $x_t = c_t$. On banking activity, the existing literature assumes that the cost of intermediation *in fixed proportion to household savings*. In the context of our model, this requires $s^i = s$ i.e. firms are all the same with respect to intermediation costs, or - equivalently - there is only one firm.

Finally, the relation of our model to the optimal inflation tax literature is as follows. Our modelling of household demand for intermediation services is closely related to the "transactions cost" view of the demand for money in that literature (Correia and Teles(1996), (1999)). In particular, if we define x_t as real money balances, their transactions cost function is an inversion of (3.3) to obtain h_t as a function of c_t, x_t ; then, increased real money balances reduce the labour transactions costs of consumption. The models in this literature do not allow for physical capital or taxation of capital income, or costly money, and so in this sense are more special. Nevertheless, one of our results, Proposition 3 below, is related to that literature, especially Proposition 2 of Correia and Teles(1996).

3.5. A Benchmark Indeterminacy Result

Here, we make the assumptions of the existing literature (Auerbach and Gordon(2002), Boadway and Keen(2003)), Jack(1999), Grubert and Mackie(1999)), namely: (i) that conditional demand for x_t is independent of h_t and linear in c_t i.e. $x_t = Ac_t$; (ii) only one type of firm; and (iii) a fixed consumption tax τ_t^c and a zero capital income tax, $\tau_t^r = 0$. Under these assumptions, we show that optimal taxes on financial intermediation are generally indeterminate. Note from (3.8)-(3.11) that given (i) i.e. $\phi_{ct} = 1$, and $\tau_t^r = 0$, we have:

$$\frac{\beta u_{ct}}{u_{ct-1}} = \frac{1 + \tau_t^c + A(1 + \tau_t^x)}{1 + \tau_{t-1}^c + A(1 + \tau_{t-1}^x)} \frac{1}{1 + r_t}, \quad t = 1, \dots \quad (3.16)$$

¹⁴It also has labour supply in only one period.

Moreover, from (3.12), (3.15), given only one firm:

$$r_t = F_{kt} - 1 - (1 + \tau_t^s)s, \quad t = 1, \dots \quad (3.17)$$

where $F_{kt} = F_k(m_t, k_t)$. Then (3.16) becomes

$$\frac{\beta u_{ct}}{u_{ct-1}} = \frac{1 + \tau_t^c + A(1 + \tau_t^x)}{1 + \tau_{t-1}^c + A(1 + \tau_{t-1}^x)} \frac{1}{F_{kt} - (1 + \tau_t^s)s}, \quad t = 1, \dots \quad (3.18)$$

Now say that the sequence $\{\tau_t^x, \tau_t^s\}_{t=0}^\infty$ is a *restricted optimal tax structure on financial services* if the inter-temporal allocation of consumption is left undistorted by taxes. From (3.18), this requires:

$$\frac{1 + \tau_t^c + A(1 + \tau_t^x)}{1 + \tau_{t-1}^c + A(1 + \tau_{t-1}^x)} \frac{1}{F_{kt} - (1 + \tau_t^s)s} = \frac{1}{F_{kt} - s}, \quad t = 1, \dots \quad (3.19)$$

Then two conclusions that can easily be drawn from (3.19). First, $\{\tau_t^x, \tau_t^s\}_{t=0}^\infty$ is not uniquely determined from (3.19) i.e. there is indeterminacy in the restricted optimal tax structure. The second is that of the many optimal tax combinations, $\tau_t^x = \tau_t^c$, $\tau_t^s = 0$ has the advantage that it is optimal, independently of knowledge of F_k, s and is thus administratively convenient. We can thus summarize:

Proposition 1. *In the benchmark case, with (i) conditional demand for x_t independent of h_t and linear in c_t ; (ii) only one type of firm; and (iii) a fixed consumption tax τ_t^c and zero capital income tax, $\tau_t^r = 0$, then the restricted optimal tax structure on financial services is not uniquely determined. But, a uniform tax on goods and payment services ($\tau_t^x = \tau_t^c$), and a zero tax on the spread ($\tau_t^s = 0$) is an administratively convenient restricted optimal tax structure.*

This result summarizes the findings of the existing literature, in the context of our model. It is important to emphasize that under the assumptions made by the existing literature, optimal taxes on financial intermediation are in fact *indeterminate*. This main purpose of this paper is to relax these assumptions in an empirically plausible way, and at the same time generate determinacy in the tax structure.

4. Tax Design

We take a primal approach to the tax design problem. In this approach, an optimal policy for the government is a choice of all the primal variables in the model, in this case $\{c_t, l_t, h_t, k_{t+1}, g_t, (k_t^i, m_t^i)_{i=1}^n\}_{t=0}^\infty$ to maximize utility (3.1) subject to the capital and labour

market clearing conditions (3.13), aggregate resource, and implementability constraints. We are thus assuming, following Chamley(1986), that the government can pre-commit to policy at $t = 0$. The aggregate resource constraint says that total production must equal to the sum of the uses to which that production is put:

$$c_t + \phi(c_t, h_t) + k_{t+1} + g_t + \sum_{i=1}^n s^i k_t^i = \sum_{i=1}^n F^i(k_t^i, m_t^i), \quad t = 0, 1, \dots \quad (4.1)$$

The implementability constraint ensures that the government's choices also solve the household optimization problem. First, by definition,

$$\phi_t \equiv \phi_{ct}c_t + \phi_{ht}h_t + C_t, \quad C_t = \phi_t - \phi_{ct}c_t - \phi_{ht}h_t \quad (4.2)$$

where C_t is the *overhead cost of payment services*. The most plausible case is where there is a fixed cost to payment services i.e. $C_t > 0$. Substituting (4.2) back into (3.7), we obtain:

$$\begin{aligned} & \sum_{t=0}^{\infty} p_t (c_t(1 + \tau_t^c + \phi_{ct}(1 + \tau_t^x)) + \phi_{ht}(1 + \tau_t^x)h_t + k_{t+1}) \\ &= \sum_{t=0}^{\infty} p_t (w_t(1 - l_t - h_t) - C_t(1 + \tau_t^x) + (1 + \rho_t)k_t) \end{aligned} \quad (4.3)$$

Then, using the household's first-order conditions (3.8)-(3.11) in (4.3), we finally arrive at the government's implementability constraint:

$$\sum_{t=0}^{\infty} \beta^t (u_{ct}c_t - u_{lt}(1 - l_t + \pi_t)) = 0 \quad (4.4)$$

where in (4.4), the expression:

$$\pi_t \equiv \frac{C_t}{\phi_{ht}} \quad (4.5)$$

is the overhead cost of payment services, normalized by $\phi_{ht} < 0$, and can thus be interpreted as the *virtual profit* of the household from transacting on the market. In the most plausible case where there is a fixed cost to payment services i.e. $C_t > 0$, virtual profit is negative.

So, as is standard in the primal approach to tax design, we can incorporate the implementability constraint (4.4) into the government's maximand by writing

$$W_t = u(c_t, l_t) + v(g_t) + \mu (u_{ct}c_t - u_{lt}(1 - l_t + \pi_t)) \quad (4.6)$$

where μ is the Lagrange multiplier on (4.4). If $u_{cl} \leq 0$, it is possible to show that $\mu \geq 0$ at the optimum (see Appendix). If $\mu = 0$, the revenue from profit taxation is sufficient to fund the public good, g . We will rule out this uninteresting case, and so will assume that $\mu > 0$ at the optimum in what follows.

The government's choice of primal variables must maximize $\sum_{t=0}^{\infty} \beta^t W_t$ subject to (4.1) and (4.4). The first-order conditions with respect to $c_t, l_t, h_t, k_t, g_t, k_{it}, m_{it}$ are, respectively;

$$\beta^t W_{ct} = (1 + \phi_{ct}) \zeta_t \quad (4.7)$$

$$\beta^t W_{lt} = \zeta_t^m \quad (4.8)$$

$$-\beta^t \mu u_{lt} \pi_{ht} = \zeta_t^m + \zeta_t \phi_{ht} \quad (4.9)$$

$$\zeta_t^k = \zeta_{t-1} \quad (4.10)$$

$$\beta^t v_{gt} = \zeta_t \quad (4.11)$$

$$\zeta_t F_{kt}^i = \zeta_t^k + \zeta_t s^i, \quad i = 1, \dots, n \quad (4.12)$$

$$\zeta_t F_{mt}^i = \zeta_t^m, \quad i = 1, \dots, n \quad (4.13)$$

where $\zeta_t, \zeta_t^k, \zeta_t^m$ are the multipliers on the resource, capital market, and labour market conditions at time t respectively.

Moreover, from (4.6),

$$\begin{aligned} W_{lt} &= u_{lt}(1 + \mu(1 + H_{lt})), \\ H_{lt} &= \frac{u_{lct}c_t - u_{llt}(1 - l_t + \pi_t)}{u_{lt}} \end{aligned} \quad (4.14)$$

and

$$\begin{aligned} W_{ct} &= u_{ct}(1 + \mu(1 + H_{ct})), \\ H_{ct} &= \frac{u_{cct}c_t - u_{cct}(1 - l_t + \pi_t) - u_{lt}\pi_{ct}}{u_{ct}} \end{aligned} \quad (4.15)$$

So, H_{ct} is what Atkeson, Chari and Kehoe(1999) call the general equilibrium expenditure elasticity. Note that if there are constant returns to scale, virtual profit $\pi_t \equiv 0$, and so H_{lt}, H_{ct} are reduce to standard formulae found, for example, in the primal approach to the static tax design problem (Atkinson and Stiglitz(1980)).

We begin by characterizing the tax on capital income and the spread tax, where we have a sharp result with strong intuition. It is possible to manipulate the first-order conditions to the household and government optimization problems to get (all proofs in Appendix):

Proposition 2. *At any date t , the optimal taxes τ_t^r, τ_t^s satisfy*

$$\left(1 + (1 - \tau_t^r) \left(\frac{\zeta_{t-1}}{\zeta_t} - 1 - \tau_t^s s^i\right)\right) = \frac{A_t}{A_{t-1}} \frac{\zeta_{t-1}}{\zeta_t}, \quad i = 1, \dots, n \quad (4.16)$$

where $A_t = \frac{(1+\phi_{ct})(1+\tau_t^c+\phi_{ct}(1+\tau_t^r))}{1+\mu(1+H_{ct})}$. So, if firms are homogenous in intermediation costs, ($s^i = s$, all i), then τ_t^r, τ_t^s are not uniquely determined, but if there is heterogeneity ($s^i \neq s^j$, some i, j) then the unique solution to the system (4.16) has $\tau_t^s = 0$, and in the steady state, $\tau_t^r = 0$.

So, we see that as long as intermediation costs differ across firms, the spread tax τ_t^s at any date should be zero. The intuition for this result is clear. From (4.12), (4.10), we see that at any date t ,

$$(F_{kt}^i - s^i) = \frac{\zeta_{t-1}}{\zeta_t} \implies F_{kt}^i - s^i = F_{kt}^j - s^j \quad (4.17)$$

That is, the marginal product of capital net of true intermediation costs should be equal across firms, which of course is just the condition for capital to be allocated efficiently across firms. But, condition (4.17) is generally not consistent with a non-zero spread tax when firms are heterogenous, as then from (3.12), (3.15),

$$F_{kt}^i = 1 + r_t + (1 + \tau_t^s)s^i \implies F_{kt}^i - s^i = 1 + r_t + \tau_t^s s^i$$

So, if $\tau_t^s s^i \neq \tau_t^s s^j$, (4.17) cannot hold. This is just an instance of the Diamond-Mirrlees production efficiency theorem. A tax on the spread is an intermediate tax on the allocation of capital, and given our assumptions (a full set of tax instruments, and no pure profits), this tax should be set to zero. Note also that when there is only one firm, this argument has no bite, and thus τ^s is left indeterminate, as in Proposition 1.

Finally, we see that in the steady state, $\tau_t^r = 0$. So, the celebrated result of Chamley(1986) that in the steady state, the tax on capital income is zero continues to hold in our setting. In this sense, the optimal structure of wage and capital income taxes is separable from the optimal tax on borrower-lender intermediation.

Next, we turn to characterize the tax on payment services. The first step is to characterize the total tax on final consumption, which from (3.8) is the weighted sum of τ_t^c and τ_t^x i.e. $\tau_t^c + \phi_{ct}\tau_t^x$. We can then state:

Proposition 3. *At any date t , the optimal total tax on final consumption in ad valorem form is*

$$\frac{\tau_t^c + \phi_{ct}\tau_t^x}{1 + \phi_{ct} + \tau_t^c + \phi_{ct}\tau_t^x} = \left(\frac{v_{gt} - \alpha_t}{v_{gt}}\right) \left(\frac{H_{lt} - H_{ct}}{1 + H_{lt}}\right), \quad \alpha_t = \frac{u_{lt}}{w_t} \quad (4.18)$$

Note that (4.18) is a formula for an optimal consumption tax that also occurs in the static optimal tax problem, when the primal approach is used (Atkinson and Stiglitz(1980, p377)). In particular, v_{gt} is the marginal benefit of \$1 to the government, and α_t is a measure of the marginal utility of \$1 to the household, so $\frac{v_{gt}-\alpha_t}{v_{gt}}$ is a measure of the social gain from additional taxation at the margin. But, inspection of (4.14) and (4.15) reveals that in our analysis, the H_{lt}, H_{ct} are generally different to the static case, unless $\pi_t = 0$, which occurs when there are constant returns in the conditional demand for payment services, $\kappa = 1$. Note also that the optimal tax $\tau_t^c + \phi_{ct}\tau_t^x$ on final consumption is a weighted average of two taxes on marketed goods, c_t and x_t , and thus these two separate taxes are not yet determinate.

The next result characterizes τ_t^x , and can be stated as follows¹⁵:

Proposition 4. *If household demand for payment services depends on the time input ($\phi_{ht} < 0$), any date t , the optimal ad valorem tax on payment services is*

$$\frac{\tau_t^x}{1 + \tau_t^x} = -\frac{\mu\alpha_t}{v_{gt}}\pi_{ht} \quad (4.19)$$

where π_{ht} is the marginal effect of h_t on virtual profit (4.5). But, if conditional demand for payment services is independent of the time input ($\phi_{ht} = 0$), then the optimal tax on payment services is indeterminate.

That is, generally, τ_t^x is determinate, but under the special conditions of the existing literature, when $\phi_{ht} = 0$, it is not. In the main case of interest, when τ_t^x is determinate, we see that it is *not* general equal to the right-hand side of (4.18), but is instead determined by the effect of h_t on the the virtual profit of the household, π_t . This of course, implies that in general, financial services should not be taxed at the same rate as the consumption good i.e. $\tau_t^c \neq \tau_t^x$, contrary to the claims of the existing literature.

So, how *is* τ_t^x determined? First, the sign of τ_t^x is the sign of $-\pi_{ht}$. One intuition for this is as follows. If the government imposes a positive tax on x_t , this will cause a reduction in x_t , and at a fixed level of consumption, c_t , a compensating increase in h_t . If this decreases virtual profit for the household, which is not directly taxable, this is desirable. But this last effect is measured just by $-\pi_{ht}$. Note that in the special case of constant returns of ϕ , then $\pi_t \equiv 0$, $\tau_t^x = 0$. This can be understood as an instance of the

¹⁵As noted in Section 2, Proposition 3 is related to Proposition 2 of Corriea and Teles(1996). They consider what is formally a very similar tax design problem. The main differences are; (i) Proposition 3 extends their analysis by providing an explicit formula for the optimal tax rate; (ii) they work with a different specification of (3.3), namely where h_t is the dependent variable.

Diamond-Mirrlees Theorem; if household "profit" is zero, the intermediate good, payment services, should not be taxed.

More generally, there is an analogy here with the Corlett-Hague rule, which says that goods complementary with non-taxable leisure should be taxed more heavily. An analogy can also be drawn with tax design when there are non-constant returns to scale in the production of marketed goods. In that case, it has long been known that in this situation, a deviation from aggregate production efficiency (non-taxation of intermediate goods) is optimal. For example, Stiglitz and Dasgupta(1971) show that factors of production should be taxed more heavily when used in industries where pure rent is positive and cannot be taxed at 100%. Here, the principle is similar: the factor of production, x_t , should be taxed (subsidized), if it causes h_t - profit to rise (fall).

We can now focus on the determinants of the sign of π_{ht} . We start with the specification (3.4), affine conditional demand. In this case, it is clear that $\pi_t = \frac{X}{ab}$ and thus $\pi_{ht} = 0$, implying a zero tax on payment services¹⁶. An alternative form for ϕ can be derived from the literature on optimal inflation tax, where it is often assumed that $\phi(c_t, h_t)$ is a homogenous function (Kimbrough(1986), Guidotti and Vegh(1993), Correia and Teles(1996, 1999)). In this case, without loss of generality¹⁷, we can write $\phi(c, h) = f\left(\frac{c}{h}\right) c^\kappa$, with $f(0) = 0$, $f' > 0$. This is homogeneous of degree κ , and is convex iff $f'' > 0$, so we assume these properties. It is then possible to show:

Proposition 5. Assume $\phi(c, h) = f\left(\frac{c}{h}\right) c^\kappa$.

(a) If in addition, $f\left(\frac{c}{h}\right) = \left(\frac{c}{h}\right)^\varepsilon$ is constant elasticity, then at any date t , the optimal ad valorem tax on payment services is

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{\mu\alpha_t(1 - \kappa)}{v_{gt} \varepsilon} \quad (4.20)$$

So, if there are decreasing returns to scale ($\kappa < 1$), then τ_t^x is positive, and if there are increasing returns to scale ($\kappa > 1$), then τ_t^x is negative.

(b) In the general case, at any date t , the optimal ad valorem tax on payment services is

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{\mu\alpha_t}{v_{gt}} \left(\frac{2 + \eta_t}{\varepsilon_t} - 1 \right) (1 - \kappa), \quad \varepsilon_t = \frac{f'_t c_t}{f_t} > 1, \quad \eta = \frac{f''_t c_t}{f'_t} > 0 \quad (4.21)$$

So, if $\varepsilon_t < 2 + \eta_t$, and if there are decreasing (increasing) returns to scale, then τ_t^x is positive (negative).

So, in a wide variety of cases, $\tau_t^x > 0$ if there are decreasing returns to scale. This is the plausible case, as it corresponds to some fixed-cost element of payment services.

¹⁶Note that this not a special case of constant returns, as (3.4) is not constant returns in h_t, c_t .

¹⁷This is true because $\phi\left(\frac{c}{h}, 1\right) = c^{-\kappa} \phi(c, h)$, so $\phi(c, h) = c^\kappa \phi\left(\frac{c}{h}, 1\right) \equiv c^\kappa f\left(\frac{c}{h}\right)$.

But, we cannot rule out a negative tax on payment services. So, overall, the conclusion is that while the optimal tax on payment services is likely to be positive or zero, a subsidy cannot be ruled out.

5. Many Consumption Goods

So far, we have assumed only one final consumption good for the household. With many consumption goods, the new issue is when it is optimal to have uniform commodity taxation *within* the period and what this implies for taxation of payment services. We now extend the baseline model to accommodate many consumption goods. To focus on payment services, we assume that there is only one firm, for whom the cost of savings intermediation, s , is zero. In each period, the consumer has preferences over n goods, $u(\mathbf{c}_t, l_t)$, $\mathbf{c}_t = (c_{1t}, \dots, c_{nt})$.

To keep things reasonably simple, we suppose that demand for payment services x_{it} used for purchase of good i takes the Cobb-Douglas form

$$x_{it} = \phi^i(c_{it}, h_{it}) = c_{it}^{\kappa^i + \theta^i} h_{it}^{-\theta^i}, \quad \theta^i \geq 0, \quad \kappa^i > 0, \quad i = 1, \dots, n \quad (5.1)$$

where h_{it} is the amount of time services that are used, and κ^i is the returns to scale in the conditional demand for x_{it} . This encompasses the case studied by Auerbach and Gordon(2002), who (implicitly) assume that $\theta^i = 0$ and $\kappa^i = 1$ ¹⁸.

If every consumption good has its own dedicated payment service, then Propositions 3 and 4 carry over basically unchanged. However, this is very unrealistic; in practice, households use just a small number of checking accounts, credit cards, etc. So, we focus on more realistic scenario where there is just type of one payment service, which must be taxed at the same rate in all its uses i.e. every x_{it} must be taxed at the rate τ_t^x . As we will see, this imposes additional constraints on the tax design problem. The time and present-value budget constraints are modified in the obvious way to

$$l_t = 1 - m_t - \sum_{i=1}^n h_{it} \quad (5.2)$$

and

$$\sum_{t=0}^{\infty} p_t \left(\sum_{i=0}^n c_{it} (1 + \tau_t^i) + \sum_{i=0}^n \phi^i(c_{it}, h_{it}) (1 + \tau_t^x) \right) + k_{t+1} = \sum_{t=0}^{\infty} p_t (w_t m_t + (1 + \rho_t) k_t) \quad (5.3)$$

¹⁸See page 412 of their paper, where it is stated that "time costs..do not enter into the budget constraint as written here..", implying $\theta^i = 0$, and where there is a fixed transaction cost per unit of the good consumed, requiring that transactions cost are linear in c_{it} .

respectively. In (5.3), the government now has separate commodity taxes τ_t^i for each of the n goods.

Then, the household chooses $\{(c_{it}, x_{it}, h_{it})_{i=1}^n, l_t, k_{t+1}\}_{t=1}^\infty$, to maximize $\sum_{t=0}^\infty \beta^t u(\mathbf{c}_t, l_t)$ subject to (5.1), (5.2), (5.3). It is then straightforward to set up the government's tax design problem, in particular the implementability constraint, much as in the previous Section. The main difference is that because all the x_{it} must be taxed at the same rate, from (A.16), the household's first-order condition for the choice of x_{it} , the derivatives of $\phi^i(c_{it}, h_{it})$ with respect to h_{it} , denoted ϕ_{ht}^i , must all be the same across goods, so we need to impose the additional constraints on the government that

$$\phi_{ht}^i = \tilde{\phi}_{ht}, \quad i = 1, \dots, n \quad (5.4)$$

The government can, however, choose $\tilde{\phi}_{ht}$, because this is the equivalent, in the primal problem, of choosing the tax τ_t^x . This problem can again be solved using the primal approach (see Appendix B). Note that the overall effective commodity tax on good i in period t is $\tau_t^i + \phi_{ct}^i \tau_t^x$, where ϕ_{ct}^i is the derivative of $\phi^i(c_{it}, h_{it})$ with respect to c_{it} . We can prove the following generalizations of Propositions 3 and 4:

Proposition 6. (i) *At any date t , the optimal total ad valorem tax on final consumption good i is*

$$\frac{\tau_t^i + \phi_{ct}^i \tau_t^x - (\chi_{it} - \bar{\chi}_t) \tilde{\phi}_{ht} / \zeta_t}{1 + \tau_t^i + \phi_{ct}^i (1 + \tau_t^x)} = \left(\frac{v_{gt} - \alpha_t}{v_{gt}} \right) \left(\frac{H_{lt} - H_{it}}{1 + H_{lt}} \right), \quad \alpha_t = u_{lt} / w_t \quad (5.5)$$

where H_{lt}, H_{it} are defined in (A.29), (A.30) in Appendix B, and χ_{it} are the multipliers on (5.4), and $\bar{\chi}_t = \frac{1}{n} \sum_{i=1}^n \chi_{it}$.

(ii) *At any date t , the optimal ad valorem tax on payment services is*

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{\mu \alpha_t}{v_{gt}} \frac{1}{n} \sum_{i=1}^n \frac{(1 - \kappa^i)}{\kappa^i} \frac{1}{\theta^i} - \frac{1}{n} \sum_{i=1}^n \frac{(\theta^i + 1)(\chi_{it} - \bar{\chi}_t) \tilde{\phi}_{ht}}{h_{it} w_t \zeta_t} \quad (5.6)$$

In the special case where $\frac{(1 - \kappa^i)}{\kappa^i} \frac{1}{\theta^i} = A, i = 1, \dots, n$, then

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{\mu \alpha_t}{v_{gt}} A \quad (5.7)$$

So, comparing Proposition 6 with Propositions 3 and 4, we see several qualitative differences, due to the constraints (5.4). First, the formula for the overall tax on good i , $\tau_t^i + \phi_{ct}^i \tau_t^x$, does not precisely follow an Atkinson-Stiglitz type formula, due to the additional

effect $-(\chi_{it} - \bar{\chi}_t)\tilde{\phi}_{ht}/\zeta_t$, which may be positive or negative, depending on whether the constraint $\phi_{ht}^i = \tilde{\phi}_{ht}$ binds more or less tightly than the average across commodities. Second, the formula for τ_t^x now depends on the *weighted average* of economies of scale in the conditional demand functions, i.e. $\frac{1}{n} \sum_{i=1}^n \frac{(1-\kappa^i)}{\kappa^i} \frac{1}{\theta^i}$, but unless $\frac{(1-\kappa^i)}{\kappa^i} \frac{1}{\theta^i} = A$, $i = 1, \dots, n$, there is an adjustment given by the last term in (5.6), reflecting the additional constraints (5.4).

We can now address the question of when taxation on final consumption goods is uniform, and what implications this has for the taxation of payment services. There are two straightforward special cases here. The first is where conditional demand (5.1) is constant returns i.e. $\kappa^i = 1$, $i = 1, \dots, n$. Then, $\frac{(1-\kappa^i)}{\kappa^i} \frac{1}{\theta^i} = A = 0$, and so from (5.7), $\tau_t^x = 0$, all i . Moreover, from (A.29) in Appendix B,

$$H_{it} = \frac{1}{u_{it}} \left(\sum_{j=1}^n u_{jt} c_t - u_{lit} (1 - l_t) \right) \quad (5.8)$$

Under certain well-known conditions, this H_{it} is independent of i , for example, if $u(\mathbf{c}_t, l_t) = u(\tilde{u}(\mathbf{c}_t), l_t)$, \tilde{u} homothetic. It then follows from (5.5) that the tax on the consumption good, and on final consumption overall, is uniform i.e. $\tau_t^i = \tau_t$, at a rate given by the right-hand side of (5.5).

The second special case is where $\theta^i = \kappa^i$, so that ϕ^i is independent of h_{it} , in which case, as in Proposition 2, *only* the weighted average of τ_t^x and τ_t^i is well-defined via (5.5). In this case, H_{it} is again defined in (5.8), so in this case, the left-hand side of (5.5) is constant across i if $u(\mathbf{c}_t, l_t) = u(\tilde{u}(\mathbf{c}_t), l_t)$, \tilde{u} homothetic. But then, this uniform final consumption tax can be implemented by a uniform tax on both goods and payment services, independently of the weights ϕ_{ct}^i . We can summarize as follows:

Proposition 7. *Assume $u(\mathbf{c}_t, l_t) = u(\tilde{u}(\mathbf{c}_t), l_t)$, \tilde{u} homothetic. Then if $\kappa^i = 1$, all i , $\tau_t^i = \tau_t$, $\tau_t^x = 0$. If $\theta^i = \kappa^i$, so that ϕ^i is independent of h_{it} , all i , then the ad valorem tax on final consumption,*

$$\frac{\tau_t^i + \phi_{ct}^i \tau_t^x}{1 + \tau_t^i + \phi_{ct}^i (1 + \tau_t^x)}$$

is uniform across goods, and this can be implemented by a common uniform tax across marketed goods and the payment service i.e. $\tau_t^i = \tau_t^j = \tau_t^x$, all $i, j = 1, \dots, n$.

So, we see that when there is no time input to household production, a uniform consumption tax on goods and financial services at the same rate may be *an* optimal tax structure (but not the only one). This directly generalizes the results of Auerbach and Gordon(2002). Their model, is a finite horizon version of the model of this section, where

additionally, it is assumed that there is no time cost of purchasing any of the n goods. They show that in this environment, *if there is a uniform tax already in place on the n marketed goods* i.e. $\tau_t^i = \tau_t$, then it is best to set $\tau_t^x = \tau_t$ also¹⁹. We have extended this result by explicitly deriving the optimal tax structure and showing under what conditions a common uniform tax across marketed goods and the payment service is desirable.

In particular, if there is no time cost of purchasing any of the n goods, a uniform tax on goods and the payment service is optimal under the standard conditions on preferences required for uniformity i.e. separability in goods and leisure and the goods sub-utility function homothetic. However, this result does *not* generally extend to the case where h_{it} i.e. time does substitute for payment services. For example, from Proposition 7, we see that with a Cobb-Douglas specification and constant returns, for example, the optimal tax on payment services is zero. Thus, the Auerbach and Gordon(2002) result depends crucially on lack of substitutability between time costs and payment services²⁰.

6. Extensions

6.1. Payment Services as Household Production

Our findings in Section 5 also relate to the small literature on optimal taxation with household production, in particular Kleven(2004), who adopts Becker's(1965) household production framework²¹. First, note that in our framework, the link between the level of the good i actually consumed - z_i in Becker's notation - and the amount purchased in the market, c_i , along with payment services x_i and time h_i can be written as follows, where we have dropped time subscripts for simplicity. First, the relationship (3.3), $x_i = \phi_i(c_i, h_i)$ can be inverted to give $c_i = \psi_i(x_i, h_i)$. Then, the implicit Becker production function for our model can be written

$$z_i = \min \{c_i, \psi^i(x_i, h_i)\} \quad (6.1)$$

¹⁹In particular, they show that if there is initially a wage income tax at rate τ , which is replaced by a consumption tax at equivalent rate $\tau/(1 - \tau)$, then the real equilibrium is left unchanged if and only if transaction services are also taxed at this equivalent rate.

²⁰Proposition 8 below implies that uniform taxation of goods and the payment service is also an optimal tax structure where time is an input to transactions, but where time and payment services are not substitutable i.e. Leontief technology *and* the ratio of payment services to time required is uniform across goods. So, most generally, the Auerbach-Gordon result applies in an environment where time and payment services are used in fixed proportions, along with other conditions.

²¹Other contributions include Sandmo(1990), Anderberg and Balestrino(2000), but these are less closely related.

That is, *actual* consumption is the minimum of the actual amount of market good i purchased, c_i , and the amount of c_i that can be bought using payment services x_i , and time h_i . This implies that the consumer who wishes to consume at level z_i will buy z_i units of the market good and $x_i = \phi_i(z_i, h_i)$ units of payment services, along with time input h_i .

This compares to the version of Becker's production function studied by Kleven, where each final consumption good z_i is produced via a fixed coefficients production function, where household production of c_i requires a marketed good and time in fixed proportions:

$$z_i = \min \left\{ c_i, \frac{h_i}{a_h^i} \right\} \quad (6.2)$$

and the household cares about final goods and leisure i.e. $u = u(z_1, \dots, z_n, l)$. Note the difference between our structure and Kleven's. In particular, his can formally be considered a special case of ours where the only other input to household production besides the marketed good is time h_i , and the production function is linear²². However, in practice, the interpretations are slightly different. If good i is omelettes, then c_i would be number of eggs bought in both frameworks, but in Kleven's framework, h_i would be the time needed to cook an omelette, whereas in our framework, h_i would be the amount of time used to buy the eggs.

Nevertheless, due to the formal similarity, it is helpful to compare our optimal tax rules to Kleven's. First, under some conditions on preferences, Kleven finds a very simple optimal tax structure, where relative taxes across goods just depend on the household production technology. Specifically, given (6.2), the tax on good i is just inversely proportional to σ^i , the share of the marketed good in the total (tax-inclusive) cost of one unit z_i , where

$$\sigma^i = \frac{p_i(1 + \tau_i)}{p_i(1 + \tau_i) + a_h^i w}$$

where p_i is the producer price of c_i and τ_i is the ad valorem tax on c_i (Kleven(2004), Propositions 1-3). In fact, using our primal approach, it is possible to show that in his framework, the optimal tax is²³

$$\frac{\tau_i}{1 + \tau_i} = \frac{1}{\sigma^i} \left(\frac{v_g - \alpha}{v_g} \right) \frac{(H_l - H_i)}{1 + H_l} \quad (6.3)$$

²²Formally, Kleven assumes $z_i = \min \left\{ \frac{c_i}{a_c^i}, \frac{h_i}{a_h^i} \right\}$, but by appropriate choice of units, we can set $a_c^i = 1$ w.l.o.g.

²³This can be proved along the lines of Proposition 8.

where H_i, H_l have the standard definitions as in (4.14),(4.15). This in fact generalizes Proposition 3 of Kleven(2004); the latter is the special case of (6.3) where compensated cross-price effects are zero. Then, if standard conditions for uniform taxes hold i.e. $u = u(\tilde{u}(z_1, ..z_n), l)$, where the sub-utility function is homothetic, it is well-known that $H_i = H_j$, so then τ_i is inversely proportional to σ^i .

To make the link between our results and Kleven's, assume $\psi^i(x_i, h_i) = \min \left\{ \frac{x_i}{a_x^i}, \frac{h_i}{a_h^i} \right\}$ in (6.1) i.e. payment services and time must be consumed in strict proportions. Then, our household production function (6.1) becomes

$$z_i = \min \left\{ c_i, \frac{x_i}{a_x^i}, \frac{h_i}{a_h^i} \right\} \quad (6.4)$$

Now, consider the optimal tax problem, as defined in Section 5 above, with fixed coefficients specification (6.4). Define

$$\sigma_t^i = \frac{1 + \tau_t^i + a_x^i(1 + \tau_t^x)}{1 + \tau_t^i + a_x^i(1 + \tau_t^x) + a_h^i w_t}$$

to be the share of the (tax-inclusive) cost of producing one unit of z_{it} that arises from the purchase of the marketed good *and* intermediation services. Then, it is straightforward to show:

Proposition 8. *In the case of a fixed coefficients production technology, the optimal overall tax rate on good i , $T_{it} = (\tau_t^i + a_x^i \tau_t^x)/(1 + a_x^i)$, is given by*

$$\frac{T_{it}}{1 + T_{it}} = \frac{1}{\sigma_t^i} \left(\frac{v_{gt} - \alpha_t}{v_{gt}} \right) \frac{(H_{lt} - H_{it})}{1 + H_{lt}} \quad (6.5)$$

where H_{lt}, H_{it} are defined in (A.29),(A.30) in Appendix B with $\pi_t^i \equiv 0$. So, if $u = u(\tilde{u}(c_1, ..c_n), l)$, with \tilde{u} homothetic, then $H_i = H_j$, and so, T_{it} is inversely proportional to σ_t^i .

Comparing Proposition 8 to (6.3), the formal similarity is clear. Note also that using Proposition 8, we can state two propositions that relate to Auerbach-Gordon(2002). First, under the additional assumption that the payment services used in the purchase of different commodities can be taxed at different rates $\tau_t^{x,i}$, a possible optimal tax structure is to always tax good i and the payment services used to purchase good i at the same rate i.e. $\tau_t^i = \tau_t^{x,i}$. This is because (6.5) only determines the weighted sum of τ_t^i, τ_t^x .

Second, under the conditions on preferences in Proposition 8, H_{it} is independent of i . So, assume that at a uniform tax $\tau_t^i = \tau_t^x = \tau_t$, $\sigma_t^i = \sigma_t^j$. Then, a uniform tax on both goods and the payment service is optimal. But for $\sigma_t^i = \sigma_t^j$, we require that

$$\frac{a_h^i}{1 + a_x^i} = \frac{a_h^j}{1 + a_x^j} \quad (6.6)$$

So, most generally, the Auerbach-Gordon result applies in an environment where time and payment services are used in fixed proportions, and these proportions are the same across commodities.

6.2. Endogenizing Savings Intermediation Services

We have, so far, treated the service of savings intermediation by banks in rather "black box" fashion. In particular, we have treated s_i , the amount of intermediation services per unit of capital supplied to firm i , as exogenous. However, it is clear that banks supply several different kinds of intermediation services, notably liquidity services (Diamond and Dybvig(1983)), and monitoring services (Diamond (1991), Besanko and Kanatas(1993), Holmstrom and Tirole (1997)). If these services provide externalities for the rest of the economy, then rather than

In this version of the paper, we do not attempt provide a fully microfounded version of these kinds of intermediation services, for several reasons. First, it is technically difficult to embed some explicit models of intermediation services into the dynamic optimal tax framework. For example, "endogenize" intermediation by looking at the provision of liquidity services using Diamond-Dybvig model, which is undoubtedly the pre-eminent microeconomic model of banking. While this is a topic for future work, the problem is that the Diamond-Dybvig model has a three-period dynamic structure, which is very difficult to embed within the standard infinite-horizon dynamic optimal tax model.

Second, the payoff from doing so in terms of increased insights is not really proportionate to the increased complexity. In the end, bank intermediation activity, when explicitly modelled, may (or may not) have spillovers on the rest of the economy. If there are spillovers, then the optimal tax is a Pigouvian one to internalise these spillovers. Ultimately, this is because the government can use the interest income tax to control the household's marginal rate of substitution between present and future consumption, and so any tax on intermediation services is a free instrument which can be used to internalize externalities arising from bank activity.

These general points are illustrated in a previous version of the paper, Lockwood(2010), where s_i is interpreted as the level of bank monitoring, along the lines of Holmstrom and Tirole(1997). In their framework, without monitoring, bank lending to firms is impossible, because the informational rent they demand is so high that the residual return to the bank does not cover the cost of capital. So, as monitoring is costly, the socially efficient level of monitoring is that level which just induces to bank to lend. In the case where the bank is competitive, i.e. where firm chooses the terms of the loan contract subject

to a break-even constraint for the bank, an assumption commonly made in the finance literature, this is also the equilibrium level of monitoring. In this case, savings intermediation should *not* be taxed, because doing so will violate production efficiency, as in the case with heterogeneous firms and a fixed amount of intermediation services per unit of savings. But, in the case where the bank is a monopolist i.e. it chooses the contract, it will generally choose a *higher* level of monitoring than this, in order to reduce the firm's informational rent. So, in this case, the optimal tax is a positive Pigouvian tax, set to internalize this negative externality.

7. Conclusions

This paper has considered the optimal taxation of two types of financial intermediation services (savings intermediation, and payment services) in a dynamic economy, when the government can also use wage and capital income taxes. When payment services are used in strict proportion to final consumption, and the cost of intermediation services is the same across firms, the optimal taxes on financial intermediation are generally indeterminate. But, when firms differ in the cost of intermediation services, the tax on savings intermediation should be zero. Also, when household time and payment services are substitutes in household "production" of final consumption, the optimal tax rate on payment services is determinate, and is generally different to the optimal rate on consumption goods.

We then extended the analysis to many consumption goods in each period. We show that the presence of payment services generally makes it less likely that commodity taxation will be uniform within the period. Moreover, *even if* conditions hold for commodity taxation to be uniform, this does not imply that payment services need to be taxed at the same uniform rate as commodities. These results generalize, and demonstrate the limitations of, the Auerbach-Gordon(2002) result that if goods are subject to a uniform tax, payment services should be subject to the same tax.

There are two obvious limitations of the analysis. The first is that the government is assumed to be able to precommit to a tax policy at time zero. However, even in a simpler setting without a banking sector, the characterization of the optimal time-consistent capital and labour taxes is a technically demanding exercise (see e.g. Phelan and Stacchetti (2001)) and so such an extension is certainly beyond the scope of this paper.

The second is the restriction to linear income taxation. The classic result of Atkinson and Stiglitz tells us that with non-linear income taxation, commodity taxation is

redundant, and more recently, Golosov et. al. (2003) has recently shown that this result generalizes to a dynamic economy. Their result would apply, for example, in a version of our model where households differ in skill levels, and without any financial intermediation. What would happen if we introduced financial intermediation in this environment? The results on taxation of payment services seem likely to be affected, as the government has additional degrees of freedom with which to tax the notional "profit" from household production.

8. References

- Anderberg, D. & A. Balestrino (2000), "Household Production and the Design of the Tax Structure," *International Tax and Public Finance*, 7(4), 563-584
- Atkeson, A., V.V.Chari, and P.J.Kehoe (1999), "Taxing capital income: a bad idea", *Federal Reserve bank of Minneapolis Quarterly Review*, 23: 3-17
- Atkinson, A., and J.E.Stiglitz (1980), *Lectures on Public Economics*, MIT Press
- Auerbach, A. and R.H.Gordon (2002), *American Economic Review*, 92 (Papers and Proceedings): 411-416
- Bianchi,J., and E.G.Mendoza, (2010), "Overborrowing, Financial Crises and "macroprudential" taxes", NBER Working Paper 16091
- Becker,G(1965), "A Theory of the Allocation of Time", *The Economic Journal*, 75, 493-517
- Besanko, D. and G.Kanatas (1993), "Credit market equilibrium with bank monitoring and moral hazard", *The Review of Financial Studies*, 6: 213-232
- Boadway, R. and M.Keen (2003), "Theoretical perspectives on the taxation of capital and financial services", in Patrick Honahan (ed.) *The Taxation of Financial Intermediation* (World Bank and Oxford University Press),, 31-80.
- Chamley, C., (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives", *Econometrica*, 54: 607-622
- Chia, N-C. and J.Whalley, (1999), "The tax treatment of financial intermediaries", *Journal of Money, Credit and Banking*, 31: 704-19
- de la Feria, R., and B. Lockwood (2010), "Opting for Opting In? An Evaluation of the European Commission's Proposals for Reforming VAT on Financial Services", *Fiscal Studies*, 31(2): 171-202
- Correia, I., and P.Teles (1996), "Is the Friedman rule optimal when money is an intermediate good?", *Journal of Monetary Economics*, 38: 223-244
- Correia, I., and P.Teles (1999), "The optimal inflation tax" *Review of Economic Dynamics*, 2: 325-346
- Diamond, D.W. (1991), "Monitoring and reputation: the choice between bank loans and directly placed debt", *Journal of Political Economy*, 99: 689-721
- Diamond, D.W and Dybvig, P.H, (1983), "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91: 401-19
- Ebrill, L., M. Keen, J-P. Bodin, and V. Summers (2001), *The Modern VAT* (Washington D.C.: International Monetary Fund)
- Guidotti, P. E. and Vegh, C. A., (1993), "The optimal inflation tax when money

reduces transactions costs : A reconsideration," *Journal of Monetary Economics*, 31, 189-205

Golosov, M., N.Kocherlakota, and A.Tsyvinski, (2003), "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, 70: 569-587

Grubert, H. and J.Mackie (1999), "Must financial services be taxed under a consumption tax?", *National Tax Journal*, 53: 23-40

Guidotti, P.E. and C.A. Vegh (1993), "The optimal inflation tax when money reduces transactions costs", *Journal of Monetary Economics*, 31: 189 205

Gup, B.E., and J.W.Kolari (2005), *Commercial banking: The Management of Risk*, 3rd ed. (Wiley)

Hellmann, T. F., K.C. Murdock and J.E. Stiglitz (2000), "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?" *American Economic Review*, 90: 147-165

Hoffman, L.A., S.N. Poddar and J. Whalley (1987), "Taxation of Banking Services Under a Consumption Type, Destination Basis VAT" *National Tax Journal* 40, 547-554

Huizinga, H. (2002), "Financial Services – VAT in Europe?" *Economic Policy*, 17: 499-534

Holmstrom, B. and J.Tirole (1997), "Financial intermediation, loanable funds, and the real sector", *Quarterly Journal of Economics*, 112: 663-691

International Monetary Fund (2010), *A Fair and Substantial Contribution by the Financial Sector*, Final Report for the G20

Jack, W. (1999), "The treatment of financial services under a broad-based consumption tax", *National Tax Journal*, 53: 841-51

Keen, M. (2010), "Taxing and Regulating Banks", unpublished paper, IMF

Kimbrough, K. (1986), The optimum quantity of money rule in the theory of public finance, *Journal of Monetary Economics*, 18: 277-284.

Kleven, H.(2004), "Optimum taxation and the allocation of time", *Journal of Public Economics*, 88(3–4), 545–557

Lockwood, B, (2010), "How Should Financial Intermediation Services be Taxed?," CEPR Discussion Papers 8122

Phelan, C. and E.Stacchetti (2001), "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, 69: 1491-1518

Poddar, S.N., and M. English (1997), "Taxation of Financial Services Under a Value-Added Tax: Applying the Cash-Flow Approach" *National Tax Journal*, 50: 89-111

Sandmo,A. (1990), "Tax Distortions and Household Production", *Oxford Economic Papers*, 42, 78-90

Stiglitz, J. E., and P. Dasgupta (1971), "Differential Taxation, Public Goods, and Economic Efficiency" *The Review of Economic Studies*, 38, 151-174.

Zee, H.H. (2005), "A New Approach to Taxing Financial Intermediation Services Under a Value-Added Tax" *National Tax Journal* 53(1): 77-92

A. Appendix

A.1. Proofs of Propositions

Proof that $\mu \geq 0$. Suppose to the contrary that $\mu < 0$ at the optimum. Then, from the properties of $u(c, l)$, $H_{lt} > 0$ from (4.14). So, from (4.14), $W_{lt} < u_t$. But from (4.8),(4.11), (4.13), (3.12):

$$\beta^t W_{lt} = \zeta_t^m = \zeta_t F_{mt}^i = \beta^t v_{gt} w_t$$

So, combining the two, we see that

$$v_{gt} < u_{lt}/w_t \quad (\text{A.1})$$

But, (A.1) says that utility could be increased if 1\$ of spending on the public good were returned to the household as a lump-sum, contradicting the optimality of the policy. \square

Proof of Proposition 2. From (4.7), (4.15), we get

$$\frac{\zeta_{t-1}}{\zeta_t} = \frac{1 + \phi_{ct}}{1 + \phi_{ct-1}} \frac{\beta^{t-1} W_{ct-1}}{\beta^t W_{ct}} = \frac{B_{t-1} u_{ct-1}}{B_t \beta u_{ct}} \quad (\text{A.2})$$

where $B_t = \frac{1 + \phi_{ct}}{1 + \mu(1 + H_{ct})}$. Next, from from (4.10),(4.12),

$$F_{kt}^i - s^i = \frac{\zeta_t^k}{\zeta_t} = \frac{\zeta_{t-1}}{\zeta_t} \quad (\text{A.3})$$

So, combining (A.2) and (A.3), we get

$$\frac{u_{ct-1}}{u_{ct}} = \frac{B_t}{B_{t-1}} \beta (F_{kt}^i - s^i) \quad (\text{A.4})$$

Next, using (3.8), (3.11), $\rho_t = (1 - \tau_t^r)r_t$, and (3.15), we get:

$$\frac{u_{ct-1}}{u_{ct}} = \frac{C_{t-1}}{C_t} \beta (1 + (1 - \tau_t^r)r_t) = \frac{C_{t-1}}{C_t} \beta (1 + (1 - \tau_t^r) (F_{kt}^i - 1 - (1 + \tau_t^s)s^i)) \quad (\text{A.5})$$

where $C_t = 1 + \tau_t^c + \phi_{ct}(1 + \tau_t^x)$. Combining (A.4), A.5), and eliminating $\frac{u_{ct}}{u_{ct+1}}$, we get that:

$$(1 + (1 - \tau_t^r) (F_{kt}^i - 1 - (1 + \tau_t^s)s^i)) = \frac{A_t}{A_{t-1}} (F_{kt}^i - s^i), \quad i = 1, ..n \quad (\text{A.6})$$

where $A_t = B_t.C_t$. Finally, using (A.3) to substitute $F_{kt}^i - s^i$ by $\frac{\zeta_{t-1}}{\zeta_t}$ in (A.6), we get (4.16) as required. If $n = 1$, (4.16) is a single condition and thus τ_t^r, τ_t^s are not uniquely determined. If $n > 1$, (4.16) comprises a system of $n > 1$ equations, and it is easy to

verify that $\tau_t^r = 1 - \frac{A_t \zeta_{t-1} - 1}{\frac{A_{t-1} \zeta_t}{\zeta_{t-1} - 1}}$, $\tau_t^s = 0$ is the unique solution to this system. So, $\tau^r = 0$ is a solution in the steady state. \square

Proof of Proposition 3. From (4.7), (4.8), (4.13),(4.14),(4.15),(3.12), we have

$$\frac{W_{ct}}{W_{lt}} = \frac{u_{ct}}{u_{lt}} \frac{1 + \mu(1 + H_{ct})}{1 + \mu(1 + H_{lt})} = \frac{1 + \phi_{ct}}{w_t} \quad (\text{A.7})$$

And, from (3.8),(3.9):

$$\frac{u_{ct}}{u_{lt}} = \frac{1 + \tau_t^c + \phi_{ct}(1 + \tau_t^x)}{w_t} \quad (\text{A.8})$$

So, combining (A.7), (A.8) we get:

$$\frac{\tau_t^c + \phi_{ct}\tau_t^x}{1 + \tau_t^c + \phi_{ct}(1 + \tau_t^x)} = \frac{\mu(H_{lt} - H_{ct})}{1 + \mu(1 + H_{lt})} \quad (\text{A.9})$$

Also, from (4.8),(4.11),(4.13),(4.14),(3.12) we have:

$$\begin{aligned} u_{lt}(1 + \mu(1 + H_{lt})) &= F_{mt}^i v_{gt} = w_t v_{gt} \\ \implies \mu &= \frac{1}{1 + H_{lt}} \frac{v_{gt} - \alpha_t}{\alpha_t}, \quad \alpha_t = \frac{u_{lt}}{w_t} \end{aligned} \quad (\text{A.10})$$

Combining (A.9),(A.10) to eliminate μ , and rearranging, we get (4.18) as required. \square

Proof of Proposition 4. From (3.10), we have

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{w_t + \phi_{ht}}{w_t} \quad (\text{A.11})$$

And from (4.9), (4.13), (3.12), we get:

$$-\frac{\beta^t}{\zeta_t} \mu \frac{u_{lt}}{w_t} \pi_{ht} = \frac{F_{mt}^i + \phi_{ht}}{w_t} = \frac{w_t + \phi_{ht}}{w_t} \quad (\text{A.12})$$

But then, combining (A.11),(A.12) and using (4.11) and $\alpha_t = \frac{u_{lt}}{w_t}$ we get

$$-\frac{\mu \alpha_t}{v_{gt}} \pi_{ht} = \frac{\tau_t^x}{1 + \tau_t^x}$$

as required. \square

Proof of Proposition 5. Drop the "t" subscripts for convenience. As ϕ is homogenous of degree κ ,

$$\pi(c, h) = \frac{\phi(1 - \kappa)}{\phi_h} = -\frac{f\left(\frac{c}{h}\right) h^2}{f'\left(\frac{c}{h}\right) c} (1 - \kappa)$$

Then, differentiating, and cancelling terms,

$$\pi_h = 1 - \kappa - \frac{2f}{f' \frac{c}{h}} (1 - \kappa) - \frac{f f''}{(f')^2} (1 - \kappa)$$

So,

$$-\pi_h = (1 - \kappa) \left(-1 + \frac{2}{\varepsilon} + \frac{\eta}{\varepsilon} \right) \quad (\text{A.13})$$

Substituting (A.13) into (4.19), we obtain (4.21), as required. Note that $f', f'' > 0$ implies $\eta > 0$, and from convexity of f , $f(x) < f'(x)x$, all x , so $\varepsilon > 1$. Finally, if $f\left(\frac{c}{h}\right) = \left(\frac{c}{h}\right)^\varepsilon$, it is easy to compute that $\frac{2+\eta}{\varepsilon} - 1 = \frac{1}{\varepsilon}$, so (4.20) then follows from (4.21). \square

A.2. The Many-Good Case

The household chooses $\{c_{1t}, \dots, c_{nt}, h_{1t}, \dots, h_{nt}, l_t, k_{t+1}\}_{t=0}^\infty$ to maximize $\sum_{t=0}^\infty \beta^t u(c_t, l_t)$ subject to:

$$\sum_{t=0}^\infty p_t \left(\sum_{i=1}^n (c_{it}(1 + \tau_t^i) + \phi^i(c_{it}, h_{it})(1 + \tau_t^x)) + k_{t+1} \right) = \sum_{t=0}^\infty p_t w_t (1 - l_t - \sum_{i=1}^n h_{it}) + (1 + \rho_t) k_t \quad (\text{A.14})$$

The FOC with respect to $c_{it}, h_{it}, l_t, k_{t+1}$ respectively are:

$$\beta^t u_{it} = p_t \lambda (1 + \tau_t^i + \phi_{ct}^i (1 + \tau_t^x)), \quad i = 1, \dots, n \quad (\text{A.15})$$

$$-\phi_{ht}^i (1 + \tau_t^x) = w_t, \quad i = 1, \dots, n \quad (\text{A.16})$$

$$\beta^t u_{lt} = \lambda w_t p_t \quad (\text{A.17})$$

$$p_t = (1 + \rho_{t+1}) p_{t+1} \quad (\text{A.18})$$

where u_{it} denotes the derivative of u with respect to c_{it} . Combination of (A.15-A.18) with (A.14) gives the implementability constraint for government. This can be derived following the same steps as in the one-good case, by substituting (A.15-A.18) and

$$\phi_t^i \equiv \phi_{ct}^i c_t + \phi_{ht}^i h_{it} + C_t^i, \quad C_t^i = \phi_t^i - \phi_{ct}^i c_{it} - \phi_{ht}^i h_{it} \quad (\text{A.19})$$

in (A.14) to give:

$$\sum_{t=0}^\infty \beta^t \left(-u_{lt} (1 - l_t) + \sum_{i=1}^n (u_{it} c_{it} - u_{lt} \pi_t^i) \right) = 0$$

where

$$\pi_t^i = \frac{C_t^i}{\phi_{ht}^i} \quad (\text{A.20})$$

is the overhead cost of payment services, normalized by $\phi_{ht}^i < 0$, and can thus be interpreted as the *virtual profit* of the household from transacting on the market. In the most plausible case where there is a fixed cost to payment services i.e. $C_t > 0$, virtual profit is negative.

So, the government chooses $\left\{ c_{1t}, \dots, c_{nt}, h_{1t}, \dots, h_{nt}, l_t, k_{t+1}, g_t, \tilde{\phi}_{ht} \right\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t \left(u(c_{1t}, \dots, c_{nt}, l_t) + v(g_{1t}, \dots, g_{nt}) + \mu \left(\sum_{i=1}^n (u_{it}c_t - u_{lt}\pi_t^i) - u_{lt}(1 - l_t) \right) \right)$$

subject to the resource constraints

$$(c_{it} + \phi^i(c_{it}, h_{it})) + g_{it} + k_{t+1} = F(1 - l_t - \sum_{i=1}^n h_{it}, k_t) \quad (\text{A.21})$$

and the constraints

$$\phi_{ht}^i = \tilde{\phi}_{ht}, \quad i = 1, \dots, n \quad (\text{A.22})$$

The FOC for this problem are :

$$\beta^t W_{lt} = F_{mt} \zeta_t \quad (\text{A.23})$$

$$\beta^t W_{it} = (1 + \phi_{ct}^i) \zeta_t + \chi_{it} \phi_{ht}^i, \quad i = 1, \dots, n \quad (\text{A.24})$$

$$-\beta^t \mu u_{lt} \pi_{ht}^i + \chi_{it} \phi_{hht}^i = (F_{mt} + \phi_{ht}^i) \zeta_t, \quad i = 1, \dots, n \quad (\text{A.25})$$

$$\zeta_{t-1} = F_{kt} \zeta_t \quad (\text{A.26})$$

$$\beta^t v_{gt} = \zeta_t \quad (\text{A.27})$$

$$\sum_{i=1}^n \chi_{it} = 0 \quad (\text{A.28})$$

where ζ_t, χ_{it} are the multipliers on (A.44), (A.22) respectively.

Proof of Proposition 6. (i) Define H_{it} and H_{lt} implicitly by

$$u_{it} H_{it} = \sum_{j=1}^n u_{jit} c_{jt} - u_{lt} \pi_{it}^i - u_{lit} \left(1 - l_t + \sum_{j=1}^n \pi_t^j \right) \quad (\text{A.29})$$

$$u_{lt} H_{lt} = \sum_{i=1}^n (u_{ilt} c_t - u_{llt} \pi_{it}^i) - u_{llt} (1 - l_t) \quad (\text{A.30})$$

where u_{jit} etc. denote cross-partials. Then from (A.23), (A.24), (3.12) we have:

$$\frac{W_{it}}{W_{lt}} = \frac{u_{it} (1 + \mu(1 + H_{it}))}{u_{lt} (1 + \mu(1 + H_{lt}))} = \frac{1 + \phi_{ct}^i + \chi_{it} \tilde{\phi}_{ht} / \zeta_t}{w_t} \quad (\text{A.31})$$

And, from (A.15),(A.17):

$$\frac{u_{it}}{w_t} = \frac{1 + \tau_t^i + \phi_{ct}^i(1 + \tau_t^x)}{w_t} \quad (\text{A.32})$$

So, combining (A.49), (A.50), we get:

$$\frac{\tau_t^i + \phi_{ct}^i \tau_t^x - \chi_{it} \tilde{\phi}_{ht} / \zeta_t}{1 + \tau_t^i + \phi_{ct}^i(1 + \tau_t^x)} = \frac{\mu(H_{lt} - H_{it})}{1 + \mu(1 + H_{lt})} \quad (\text{A.33})$$

Then, using (A.10) to substitute for μ in (A.51), which still applies in this case, and using (A.28), we get (5.5) as required.

(ii) From (A.16), (A.22), and $F_{mt} = w_t$, we have

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{w_t + \tilde{\phi}_{ht}}{w_t} \quad (\text{A.34})$$

and from (A.25), (A.27), and $F_{mt} = w_t$, we get:

$$-\frac{\alpha_t}{v_{gt}} \mu \pi_{ht}^i + \frac{\chi_{it} \phi_{hht}^i}{w_t \zeta_t} = \frac{\tau_t^x}{1 + \tau_t^x}, \quad i = 1, \dots, n \quad (\text{A.35})$$

Moreover, from (A.20), we get

$$\pi_{ht}^i = \frac{(\kappa^i - 1)}{\kappa^i} \left(\frac{\phi_t^i \phi_{hht}^i}{(\phi_{ht}^i)^2} - 1 \right) \quad (\text{A.36})$$

We can compute, using (A.22), that

$$\frac{\phi_t^i \phi_{hht}^i}{(\phi_{ht}^i)^2} = \frac{1}{\theta^i} + 1, \quad \phi_{hht}^i = -(\theta^i + 1) \frac{\tilde{\phi}_{ht}}{h_{it}} \quad (\text{A.37})$$

Substituting (A.36),(A.37) into (A.35), we get

$$\frac{\tau_t^x}{1 + \tau_t^x} = \frac{\mu \alpha_t (1 - \kappa^i)}{v_{gt}} \frac{1}{\kappa^i} \frac{1}{\theta^i} - \frac{(\theta^i + 1)(\chi_{it} - \bar{\chi}_t) \tilde{\phi}_{ht}}{h_{it} w_t \zeta_t}$$

Averaging this across all goods, and using $\bar{\chi}_t = 0$ from (A.28), we get (5.6) as required.

Finally, in the case where $\frac{(1 - \kappa^i)}{\kappa^i} \frac{1}{\theta^i} = A$ holds, we can argue as follows. Consider the *less-constrained problem* for the government where we do not impose (5.4) and so the τ_{it}^x can vary across commodities. In that case, it is easy to check, using the argument above, that

$$\frac{\tau_{it}^x}{1 + \tau_{it}^x} = \frac{\mu \alpha_t (1 - \kappa^i)}{v_{gt}} \frac{1}{\kappa^i} \frac{1}{\theta^i}, \quad i = 1, \dots, n \quad (\text{A.38})$$

So, if $\frac{(1-\kappa^i)}{\kappa^i} \frac{1}{\theta^i} = A$ holds, from (A.38),

$$\frac{\tau_{it}^x}{1 + \tau_{it}^x} = \frac{\mu \alpha_t}{v_{gt}} A, \quad i = 1, \dots, n \quad (\text{A.39})$$

in the less-constrained problem, and so (A.39) must also solve the original problem. \square

Proof of Proposition 8. From (6.4), the consumer will choose $x_i = a_x^i c_i$, $h_i = a_h^i c_i$. So, the household problem is to choose $\{c_{1t}, \dots, c_{nt}, l_t, k_{t+1}\}_{t=0}^{\infty}$ to maximize $\sum_{t=0}^{\infty} \beta^t u(c_t, l_t)$ subject to:

$$\sum_{t=0}^{\infty} p_t \left(\sum_{i=1}^n (c_{it}(1 + \tau_t^i) + a_x^i(1 + \tau_t^x) + w_t a_h^i) + k_{t+1} \right) = \sum_{t=0}^{\infty} p_t w_t (1 - l_t) + (1 + \rho_t) k_t \quad (\text{A.40})$$

The FOC with respect to $c_{it}, h_{it}, l_t, k_{t+1}$ respectively are:

$$\beta^t u_{it} = q_{it} p_t \lambda, \quad q_{it} = (1 + \tau_t^i) + a_x^i(1 + \tau_t^x) + w_t a_h^i, \quad i = 1, \dots, n \quad (\text{A.41})$$

$$\beta^t u_{lt} = \lambda w_t p_t \quad (\text{A.42})$$

$$p_t = (1 + \rho_{t+1}) p_{t+1} \quad (\text{A.43})$$

Combination of (A.41-A.43) with (A.14) gives the implementability constraint for government. This can be derived similarly to the proof of Proposition 5 to give:

$$\sum_{t=0}^{\infty} \beta^t \left(\sum_{i=1}^n u_{it} c_{it} - u_{lt} (1 - l_t) \right) = 0$$

So, the government chooses $\{c_{1t}, \dots, c_{nt}, l_t, k_{t+1}, g_t\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t \left(u(c_{1t}, \dots, c_{nt}, l_t) + v(g_{1t}, \dots, g_{nt}) + \mu \left(\sum_{i=1}^n u_{it} c_{it} - u_{lt} (1 - l_t) \right) \right)$$

subject to the resource constraints

$$c_{it}(1 + a_x^i) + g_{it} + k_{t+1} = F(1 - l_t - \sum_{i=1}^n c_{it} a_h^i, k_t) \quad (\text{A.44})$$

The FOC for this problem are :

$$\beta^t W_{lt} = F_{mt} \zeta_t \quad (\text{A.45})$$

$$\beta^t W_{it} = (1 + a_x^i + a_h^i F_{mt}) \zeta_t, \quad i = 1, \dots, n \quad (\text{A.46})$$

$$\zeta_{t-1} = F_{kt} \zeta_t \quad (\text{A.47})$$

$$\beta^t v_{gt} = \zeta_t \quad (\text{A.48})$$

where ζ_t is the multipliers on (A.44) as before.

(b) Now, define H_{it} and H_{lt} by setting $\pi_t^i \equiv 0$ in (A.29) ,(A.30). Then from (A.45), (A.46), (3.12) we have:

$$\frac{W_{it}}{W_{lt}} = \frac{u_{it} (1 + \mu(1 + H_{it}))}{u_{lt} (1 + \mu(1 + H_{lt}))} = \frac{(1 + a_x^i + a_h^i F_{mt})}{w_t} \quad (\text{A.49})$$

And, from (A.15),(A.17):

$$\frac{u_{it}}{u_{lt}} = \frac{1 + \tau_t^i + a_x^i(1 + \tau_t^x) + a_h^i w_t}{w_t} \quad (\text{A.50})$$

So, combining (A.49), (A.50), we get:

$$\frac{\tau_t^i + a_x^i \tau_t^x}{1 + \tau_t^i + a_x^i(1 + \tau_t^x) + a_h^i w_t} = \frac{\mu(H_{lt} - H_{it})}{1 + \mu(1 + H_{lt})} \quad (\text{A.51})$$

Then, using (A.10) to substitute for μ in (A.51), which still applies in this case, we get

$$\frac{\tau_t^i + a_x^i \tau_t^x}{1 + \tau_t^i + a_x^i(1 + \tau_t^x) + a_h^i w_t} = \left(\frac{v_{gt} - \alpha_t}{v_{gt}} \right) \frac{(H_{lt} - H_{it})}{1 + H_{lt}}$$

But then

$$\begin{aligned} \frac{\tau_t^i + a_x^i \tau_t^x}{1 + a_x^i + \tau_t^i + a_x^i \tau_t^x} &= \frac{1 + a_x^i + a_h^i w_t + \tau_t^i + a_x^i \tau_t^x}{1 + a_x^i + \tau_t^i + a_x^i \tau_t^x} \left(\frac{v_{gt} - \alpha_t}{v_{gt}} \right) \frac{(H_{lt} - H_{it})}{1 + H_{lt}} \\ &\implies \frac{T_{it}}{1 + T_{it}} = \frac{1}{\sigma_t^i} \left(\frac{v_{gt} - \alpha_t}{v_{gt}} \right) \frac{(H_{lt} - H_{it})}{1 + H_{lt}} \end{aligned}$$

as required. \square